Statistical properties of electoral systems: the Mexican case

G. Báez and H. Hernández-Saldaña

Laboratorio de Sistemas Dinámicos,

Universidad Autónoma Metropolitana Azcapotzalco, México D. F., México

R.A. Méndez-Sánchez

Centro de Ciencias Físicas, Universidad Nacional Autónoma de México,

A.P 48-3, 62210, Cuernavaca, Mor., México

Abstract

We study some statistical properties of the results of the Mexican elections of July 2nd, 2006. Our studies only apply for the data of the program of preliminary electoral results. We show that this program does not yield the results in a random way. Numbers that should be conserved are studied statistically in detail. The distributions of the votes obtained by the different parties are obtained. Some distributions indicate small world while other can be fitted by daisy models. We also show that the election, as a measurement processes, has an error of $\sim 2\%$. Since the difference between the two main candidates for president in this data-basis is of the order of $\sim 0.5\%$ a winner cannot be given.

Keywords: Elections, opinion polls, close elections, error analysis

In almost all the world, the democracy is a very important issue. It is also important how the advent of new technologies is changing the way in which elections in democratic countries are realized. On July 2nd, 2006, elections in Mexico were realized by the Mexican Federal Institute of Elections (IFE by its spanish acronym). That institute has one of the most advanced electoral systems. In fact, the IFE yields technical support to several emerging electoral systems in the world like Iraq and Haiti. In the last Mexican elections of July 2nd, a very atypical election happened. In the opinion polls realized few weeks before the election, two of the candidates for president were almost in tie. This situation becomes a novelty since it is the first time that the IFE need to resolve a presidential election considered a tie. As any measurement process the fact that several errors are associated to the election, this situation makes difficult to decide a winner [1, 2].

In this work we analyze one of the systems implemented by the IFE say, the program of preliminary electoral results or PREP by its acronym in spanish [3]. It is supposed that this program yields in less than 3 days a very accurate result of the elections. The PREP works with certificates that come stamped over the packets of ballots. In those certificates the authorities of each electoral cabin write the number of votes for each party, total number of votes, etc. at the end of the election. Then the authorities of each electoral cabin deliver the ballots and certificates to the capture centers.

In the next section we will show the time behavior of the results that the PREP yield and we analyze the reliability of the system. Using some simple conservation laws, in Section II, we made the error analysis of the data in the PREP file. In Section III we show and analyze the distributions of votes that for the elections of deputies, senators and president. A brief conclusion, about the error of the election by using the results of the PREP follows.

I. TIME BEHAVIOR OF THE PREP

Around 20:00 hrs of July 2nd, the program of preliminary electoral results (PREP) of IFE [3] started to display and update results each 5 minutes. At the beginning, the PREP yielded a very big difference ($\sim 7\%$) in favor of the candidate Felipe Calderón of the right party (PAN by its spanish acronym [4]). But all citizens, including scientist, were waiting a very close election due to the results of the opinion polls. The tendency showed a decrease in the votes of the right party PAN and an increase in the votes of the other main

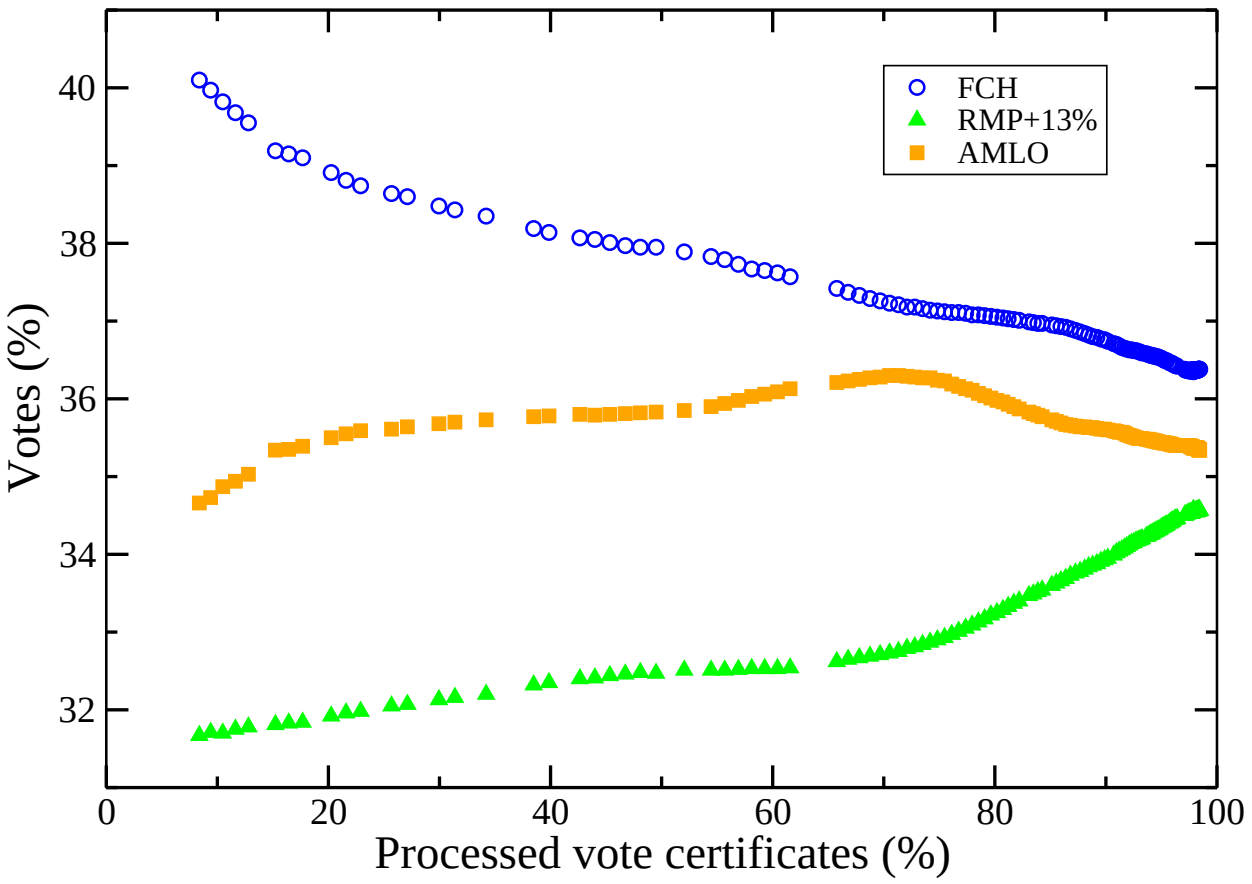

FIG. 1: Real time data given by the program of preliminary electoral results (PREP). The blue circles correspond to Felipe Calderón Hinojosa (FCH), the orange circles to Andrés Manuel López Obrador (AMLO) and the green circles to Roberto Madrazo Pintado (RMP). The data were taken from the Internet [5]. We added 13% of votes to RMP.

candidate, Andrés Manuel López Obrador of a coalition of left parties (PBT[4]). In a first sight, a crossing between the votes obtained by the two candidates was eminent. For that reason, several scientist started to capture the real time results with different methods, some captured by hand, other captured automatically with programs like *perl* [5].

In Fig. 1 we show the plot of the real time data given the PREP for the three main candidates. This plot shows roughly two tendencies for the votes obtained by the candidate of the left party. The increasing tendency changed to a decreasing tendency around 3:OO AM of 3rd July. Probably the same time in which the rural vote started to arrive to the capture centers. At the same point, Roberto Madrazo of the coalition ABM [4] started to gain votes, since one of the parties of this coalition, the PRI, has strong influence on rural

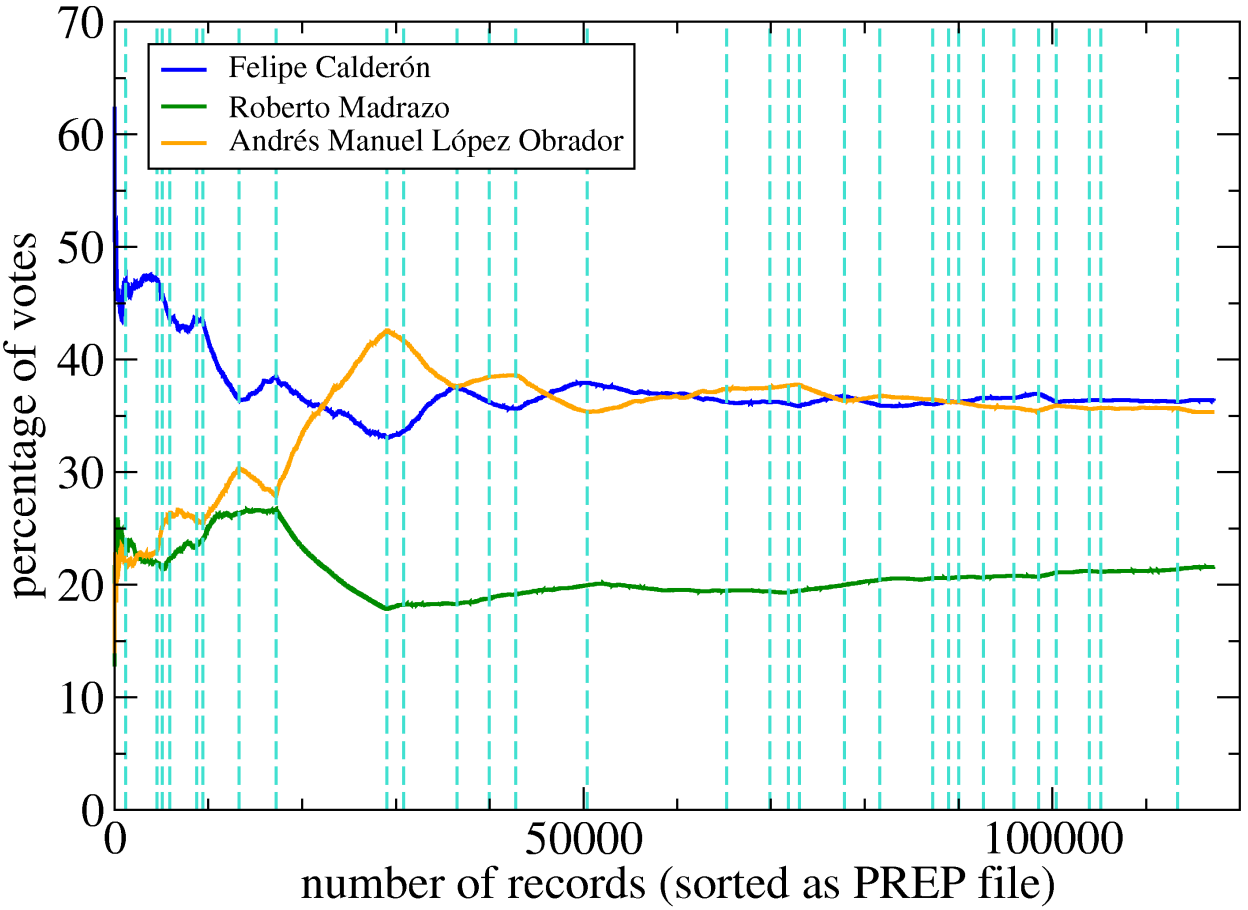

FIG. 2: Percentage of votes as a function of the number of certificate records for the three main candidates. The vertical lines separate between different states of Mexico sorted alphabetically. The big change around 20000 records corresponds to the capital of Mexico governed by the left party.

areas.

Then, no crossing between the vote percentage of the two candidates was found with the PREP in real time . Nevertheless one should notice that the real-time PREP started to display results with around 8% of the certificates processed. Although, one may think on mistrusts due to that plot, one should be careful since the PREP gives results that depend on a part of the voting which does not represent a statistical sample taken from a uniform random distribution[6, 7]. This means that, assuming a clean election, the data arrived to the capture centers from sites with better transport networks and/or a better vote counting performance. In Fig. 2 we plot the data of the PREP as ordered in their final form, by alphabetical order of the state name. As seen, the tendency of each state of Mexico is

reflected in this plot. Henceforth a clear correlation between the percentage of votes and the way we sort the vote certificates is clear. In order to broke such a correlation we sorted the certificates in a random way yielding a fast convergence to their final values, except for small fluctuations. In figure 3(a) and (b) we show two realization of such a shuffled process, one (3(a)) with the right candidate winning at the beginning and the another (3(b) with an initial advantage of the left party. Since it is not a unique way to order the sample, it is possible to sort the data in a way that any of the three main candidates has an advantage for small processed votes. In 3(c) we select all the certificates where the PRI candidate won and we putted them at the beginning or the final PREP file. Clearly this candidate remains the advantage for around the 11% and then start to lost it. In 3(d) we show the result of consider at the end of the counting process (for a random shuffle) all the cabins where the leftist candidate AMLO won ($\sim 40\%$). Note that the percentage of votes in cabins where he did not win is smaller than those to RMP-PRI.

The percentage of votes as a function of the number of records for the small parties are plotted in Fig. 4. As can be seen, the number of annulled votes is more than twice the difference between the two candidates with more votes. This is a first indication that the election yields a tie. The number of annulled votes can come from citizens who want to annulate their vote or from real errors from voters. Another interesting conclusion of this figure is that the number of votes for the non-registered candidates is very large. This means that one of the independent candidates probably obtained one or more places in the chamber since, in the Mexican laws, there are deputies proportional to the number of votes obtained. Since the names of the non-registered candidates are written by hand in the ballots, there is no other option that read and count all ballots with non-registered candidates. From Figs. 2 and 4 one can see directly that there are strong correlations between the parties that obtained large and small number of votes. It probably means that the voters can be represented by two parties at the same time. It also means that the small parties generate a distraction for the voters. This distraction becomes important in very close elections. The percentage of votes for the senators and deputies show a similar behavior to that of the president candidates with the only exception of the Nueva Alianza party. It shows a number of votes approximately 4 times larger for the chambers positions than those for president.

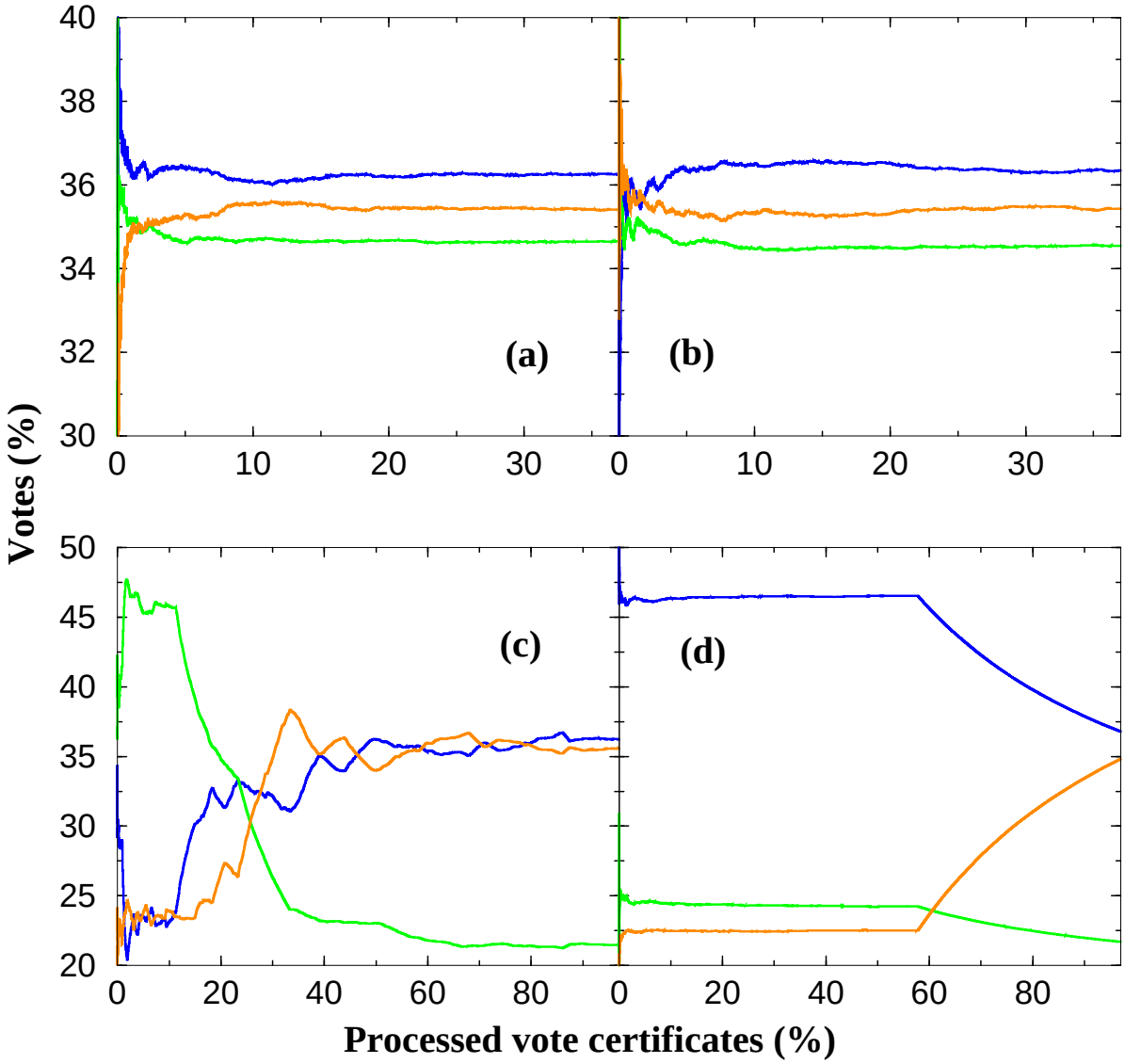

FIG. 3: (Color on line) Percentage of votes for the three main candidates as a function of processed vote certificates (%) but suffling the results in four different ways. In blue lines the percentage of votes for Felipe Calderón. The green lines correspond to Roberto Madrazo (we add a 13% to this result in (a) and (b)). In orange lines are the results for AMLO. In (a) and (b) we present two realization of a random shuffling of the PREP file. In (c) we present the results sorted as in the PREP file but with the vote certificates where Roberto Madrazo won places at the beginning. In (d), the results when vote certificates where AMLO won were placed at the end of a random shuffled file (the same realization as in (a)).

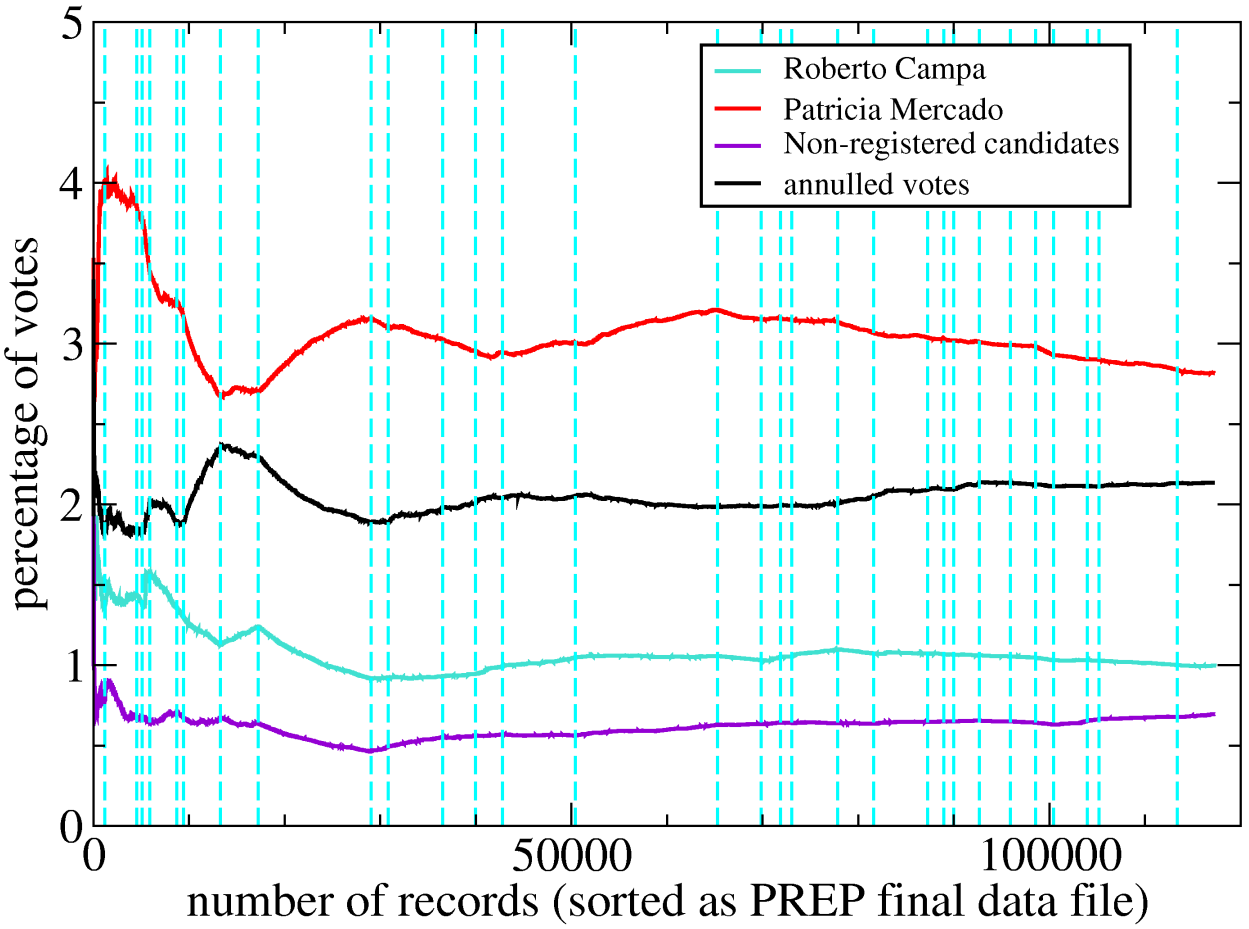

FIG. 4: Percentage of votes as a function of the number of records for the other candidates and the annulled votes. The vertical lines separate between different states of Mexico sorted alphabetically. Notice the correlation between the results of this and figure 2.

II. CONSERVATION LAWS IN ELECTIONS

In the cabin certificates became written several data: Total number of received ballots at the electoral process beginning (B_r). Number of remaining (not used) ballots (B_s). Number of voters (V). Number of deposited ballots in each cabin (B_d). And the number of votes for each party/candidate (V_i, i = PAN, PRI, PRD, Nueva Alianza,Alternativa, non-registered candidates and annulled votes). Based on this information, there are some conservation laws that can be checked in the election. For instance, one can check the lost or appearance of ballots. In particular we can study the total number of ballots minus the number of remaining ballots minus the number of voters per cabin,i.e., $B_r - (B_s + V)$. In principle this number should be zero but, as seen in Fig. 5, the distribution of this quantity

is peaked around zero but is not the expected $\delta(x)$. Data in the positive axis mean lost ballots. Data in the negative axis indicate appearance of ballots. Zooms in different parts of this distribution show the following facts. (1) The PREP conservates this number for only $\sim 45\%$ of the cabins. This result is very unfortunate for IFE since it says that the PREP is reliable only less than 50%. (2) The distribution is not completely symmetric. In particular the peak around -250 is higher than the peak at 250. (3) There are inconsistencies of more than 150 votes in several cabins. Those results are non-sense. (4) There are peaks in $\pm 10, \pm 20, \ldots$ and also in $\pm 100, \pm 200, \ldots$. Those peaks are related with capture typos. (5) The left peak of the distribution shows a different behavior for senators than deputies and president. This result cannot be understood statistically since all certificates (for president, senators and deputies) should be very similar to each other since are captured in the same way. *This means that the capture of the data was different for three similar processes.* (6) The distribution between 10 and 150 decays as a power law.

Apart from capture errors and annulled votes, other sources of errors exist. One is related to the persons who could not vote. A conservative estimation yields more than 2.5×10^5 voters. This estimation was made considering the special electoral cabins for citizens that are far from their designated electoral cabin. This number is similar to the difference between the votes obtained by the two main candidates in this election. Also, since in some states of Mexico two elections take place at the same time, errors can occur when people take less ballots, i.e., the ballots for the local election only. The opposite is also possible. Although there are errors in the counting by humans, they are diminished since several persons count the ballots several times. Up to our knowledge, the evaluation of this error has not been realized. One should notice that errors can yield also extra votes to the different candidates. Finally, other errors (random or systematic) are not taken into account.

In Figs. 6 and 7 we test other conservation laws, by ploting the distributions of difference between the total ballots received B_r minus the sum of the ballots remained B_s plus the ballots deposited in urns B_d by cabin, and the difference between the total ballots received B_r minus the sum of the ballots remained B_s plus the sum of votes obtained by each political party, included null votes $\sum_i V_i$; respectively. Those distributions should give also a δ function in the ideal case. But, as seen, one is very asymmetric. In the insets we show that typos are also present and that power laws appear.

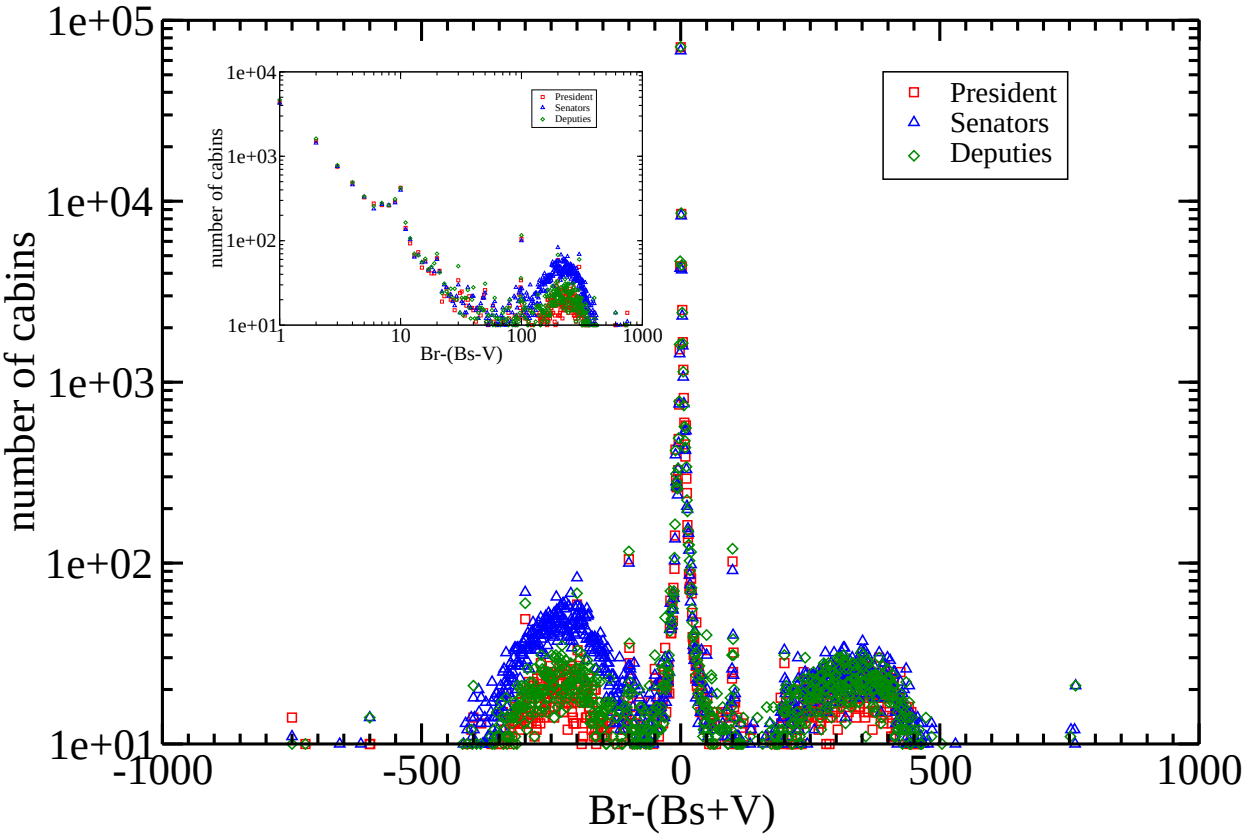

FIG. 5: Probability distribution of the difference between the total ballots received B_r minus the sum of the ballots remained B_s and the total number of voters V by cabin. The negative values in the horizontal axis means more votes than ballots and positive values means lost of ballots, but in both cases there are no conservation of the the total number of ballots received B_r in each cabin. The inset shows the left branch of the distribution in log-log scale. This shows a decay as a power law. Notice the several sharp peaks of the distribution along the horizontal axis between the values 10 to 100. These peaks show the *capture mistakes* made by humans. We noticed also that the probability distributions of the senators show a strange behavior related with the corresponding probability distributions for president and deputies in the left branch.

III. DISTRIBUTIONS OF THE VOTES

The histograms of the number of electoral cabins that have a certain number of votes is given in Fig. 8 (a). As seen the histogram for the votes of Roberto Madrazo varies very slowly (in Fig. 9 appears a fitting and below their explanation). The tail of this distribution looks exponential. Probably the form of this distribution represents the corporative vote.

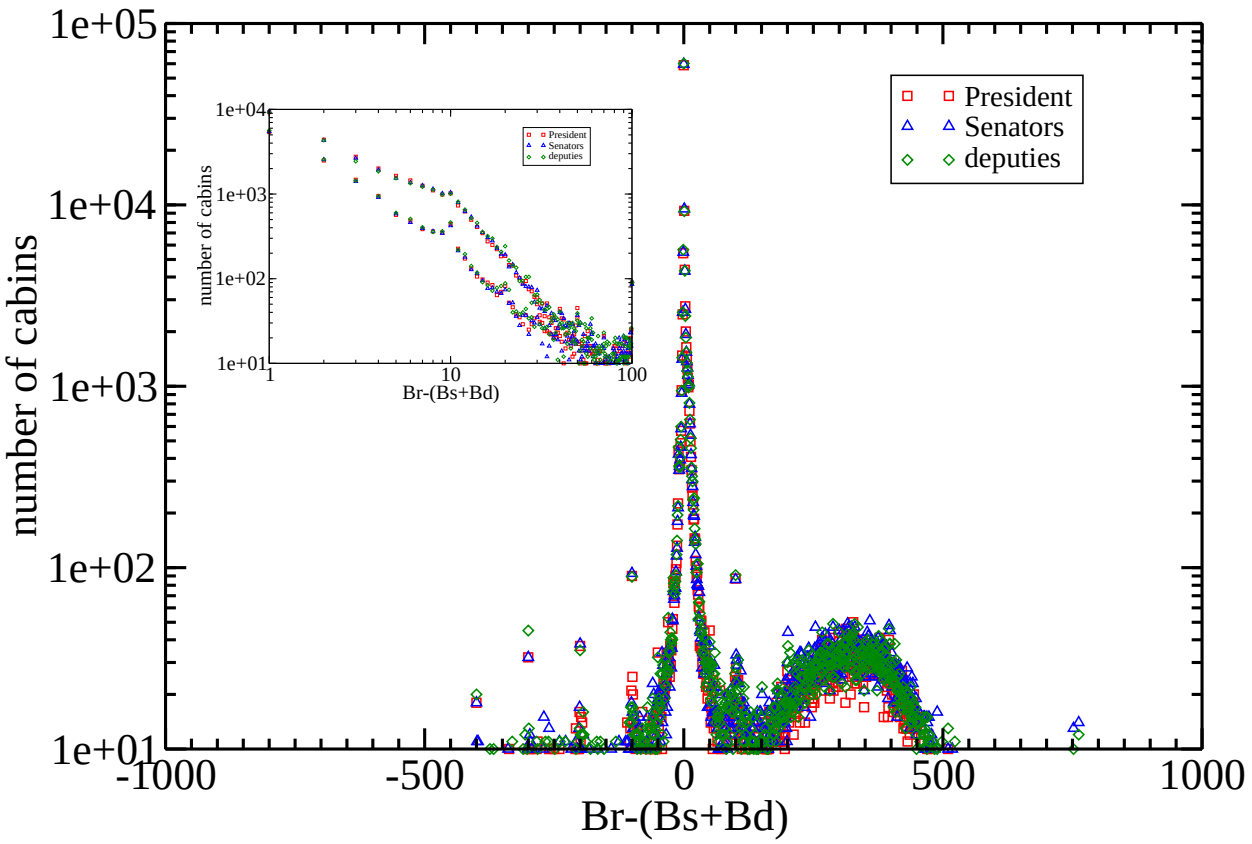

FIG. 6: Probability distribution of the difference between the total ballots received B_r minus the sum of the ballots remained B_s plus the ballots deposited in urns B_d by cabin. Notice that the asymmetry between the both branches for the three distributions is similar, but it is different from that in previous figure. Again the left branch of this probability distributions shows the sharp peaks associated to *capture mistakes*. The inset shows both branches of the distributions in log-log scale.

This is not the case for the distributions of the two main candidates. The distribution of the votes for Felipe Calderón shows a very different behavior for electoral cabins with less than ≈ 40 votes since it starts flat. The distribution of Andrés Manuel López Obrador is also very strange. It shows three different regimes. It appears like a distribution in which realizations between 60 and 300 votes are missing. This can be due to two reasons, the first one is that the data were manipulated, the second implies that the distribution of the votes for Andrés Manuel López Obrador is composed by two or more distributions [8]. This is in great contrast with the distributions of the votes for Felipe Calderón and Roberto Madrazo which vary very slowly in the same interval of votes. The distributions for the senators and

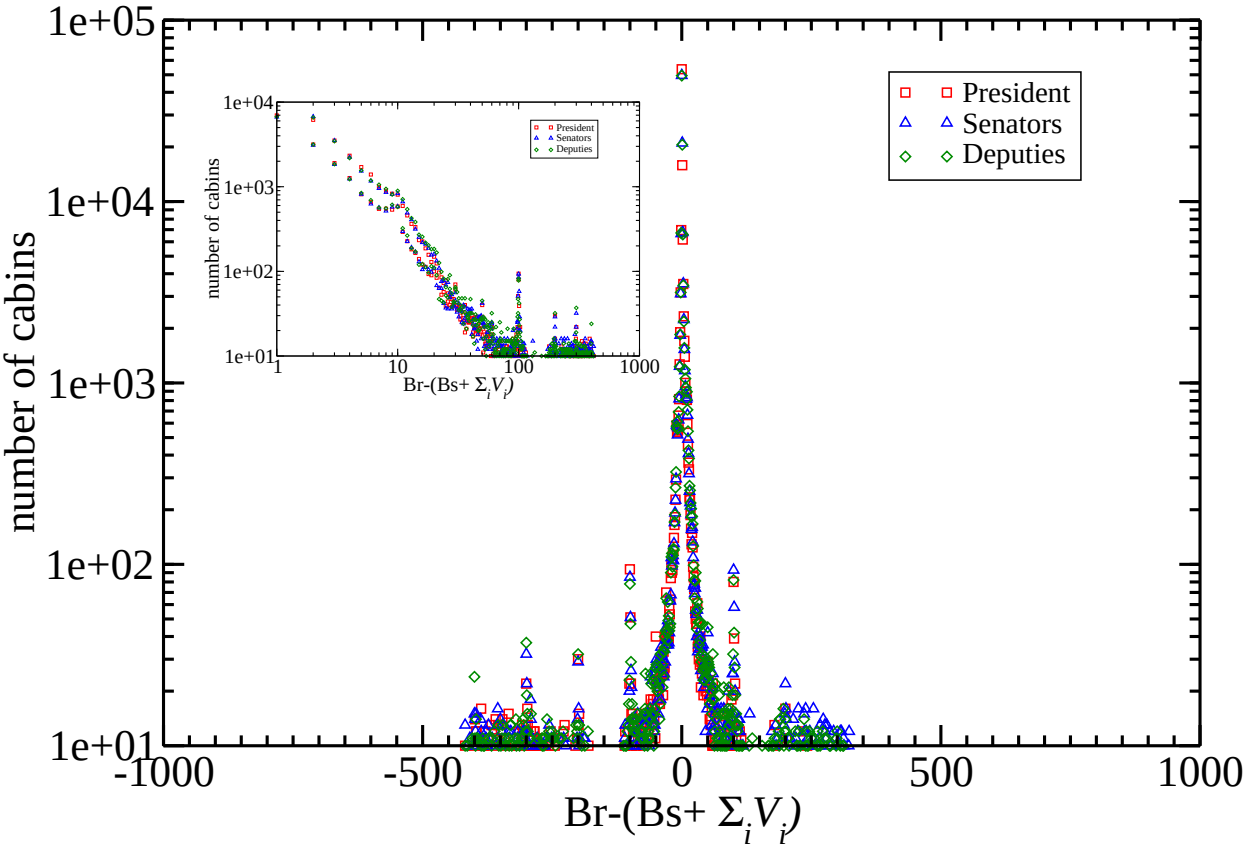

FIG. 7: Probability distribution for difference between the total ballots received B_r minus the sum of the ballots remained B_s plus the sum of votes obtained by each political party, included null votes $\sum_i V_i$. The mean of the right and the left branches of the distribution is similar to the figure 5. Notice that the three distributions are similar. The inset shows both branches of the distributions in log-log scale.

deputies present similar behavior as seen in Fig. 8 (b) and (c).

The histograms for the parties with small number of votes are given in Fig. 8 (d). As seen, all histograms have a similar behavior between them, except at small number of votes. All them are shifted power-laws, except for Roberto Campa. Those results can be explained with several models [10] on cluster growth in complex networks, for instance, and appear in other electoral processes [9]. A research on this lines is in progress [11]. We found an inconsistency in the tail of the annulled votes. This distribution shows several electoral cabins with more that 100 annulled votes. The probability of having such results is negligible. Then those results are not statistical.

As a final remark, we return to the RMP-PRI case. The distribution of votes for this

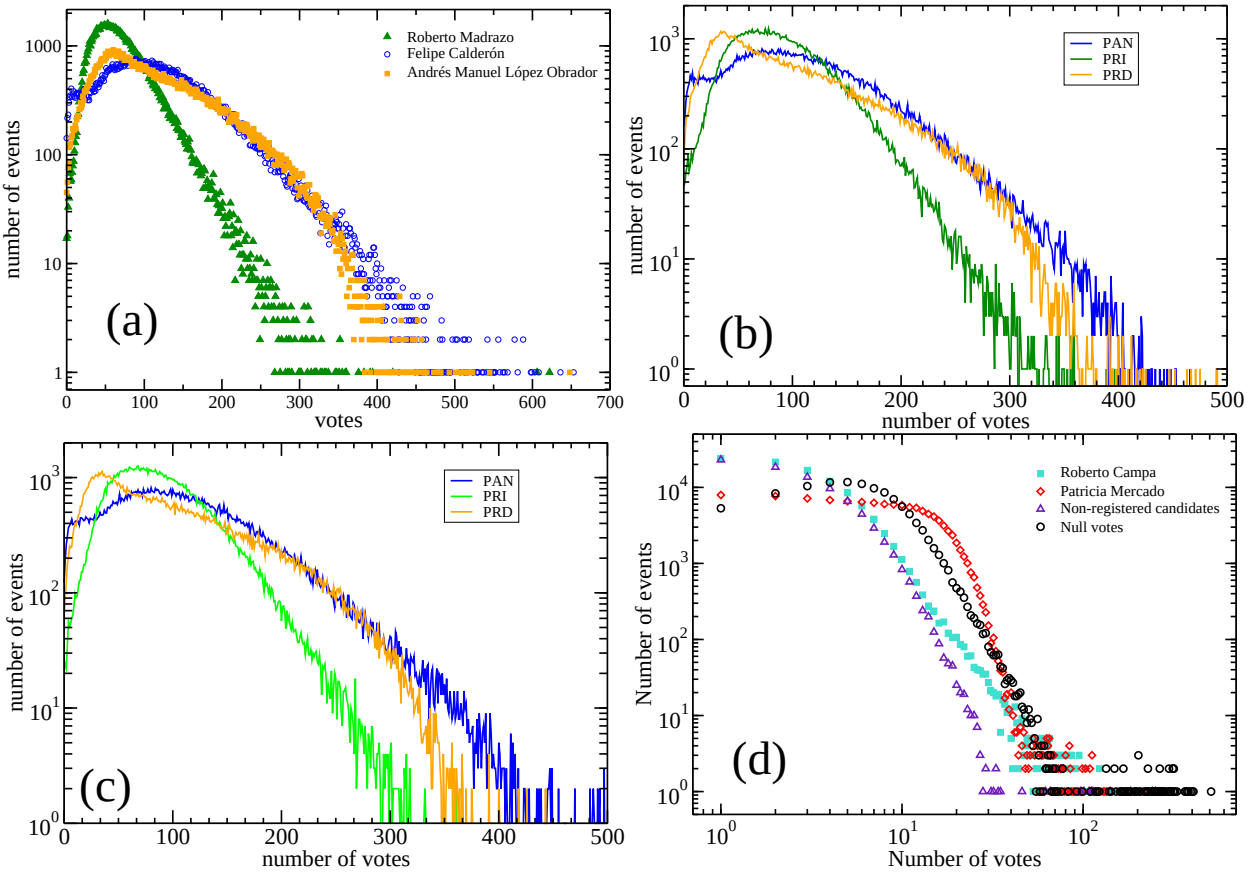

FIG. 8: Histograms showing the number of electoral cabins that obtained certain number of votes for the three main parties, (a) president, (b) deputies, (c) senators. In (d) the results obtained for president by the small parties are given. Note that the PRI distribution have an exponential tail while the distribution of the small parties show a shifted power-law.

party is clearly of a statistical nature and it is tempting to propose an analytical fit. However the lack of a straight way to obtain their natural average density makes difficult to do the usual unfolding to obtain the normalized sequence, both in its area and in its average, in order to analyze the fluctuation properties. As an anzatz, we select one realization of the randomly shuffled sequence and generate a cumulative votes density. We then adjust a 4th degree polynomial on windows of 300 and 3000 cabin certificates obtaining a reliable average density of events for the RMP-PRI votes counting. In both cases the results were similar. On this unfolded sequence of votes we calculate their nearest neighbor distribution. A larger and cautious analysis of the distribution is in progress but this result have a good agreement

with a daisy model [12] distribution

$$P_r(n,s) = \frac{(r+1)^{(r+1)n}}{\Gamma([r+1]n)} s^{(r+1)n-1} \exp[-(r+1)s]. \quad (1)$$

where n corresponds to the nth neighbor spacing distribution and r to the kind or rank of the daisy family. This model is the result of removing every $(r+1)$th level from a random sequence. In Fig. 9 we present the daisy model of rank 2 and 3 for the nearest neighbor case($n = 1$) and their comparison with the unfolded PRI distribution. The former daisy model adjust the tail and the later the central part. Since we broke correlations larger than the mean value we can expect that only the statistics around the mean is correct. Note that no fitting parameter was used, in contrast to consider a Brody function. One should remark that the daisy models are derived of a single poissonian distribution.

IV. CONCLUSIONS

To conclude, we have studied the statistical properties of the Mexican elections from the program of preliminary electoral results. We have shown that the appearence of the data in time are not statistical. Evidence of correlations between parties with large number of votes and small number of votes are evident. There is also evidence of correlation with the annuled votes. Quantities that shold be conserved where studied. Typos in capture of the data are evident. We also have obtained the distributions of votes of the different parties. In particular the distribution of the party that was in power in Mexico during more than 70 years behaves smothly. Daisy models of 2nd and 3rd rank seem to fit different parts of the measured distribution. In contrast the distributions of the parties with more votes are atypical while the distributions of small parties follow power laws. This is an evidence of small world processes. A heuristic evaluation of the error supports that it is around 8 times larger than the difference between the two main candidates. The sentence, "in a democracy, the winner of the election can be decided just by one vote" is not valid. The difference between the first and second place should be larger than the error associated to the mesurement, in this case the electoral processes. The second round in very close elections is probably needed.

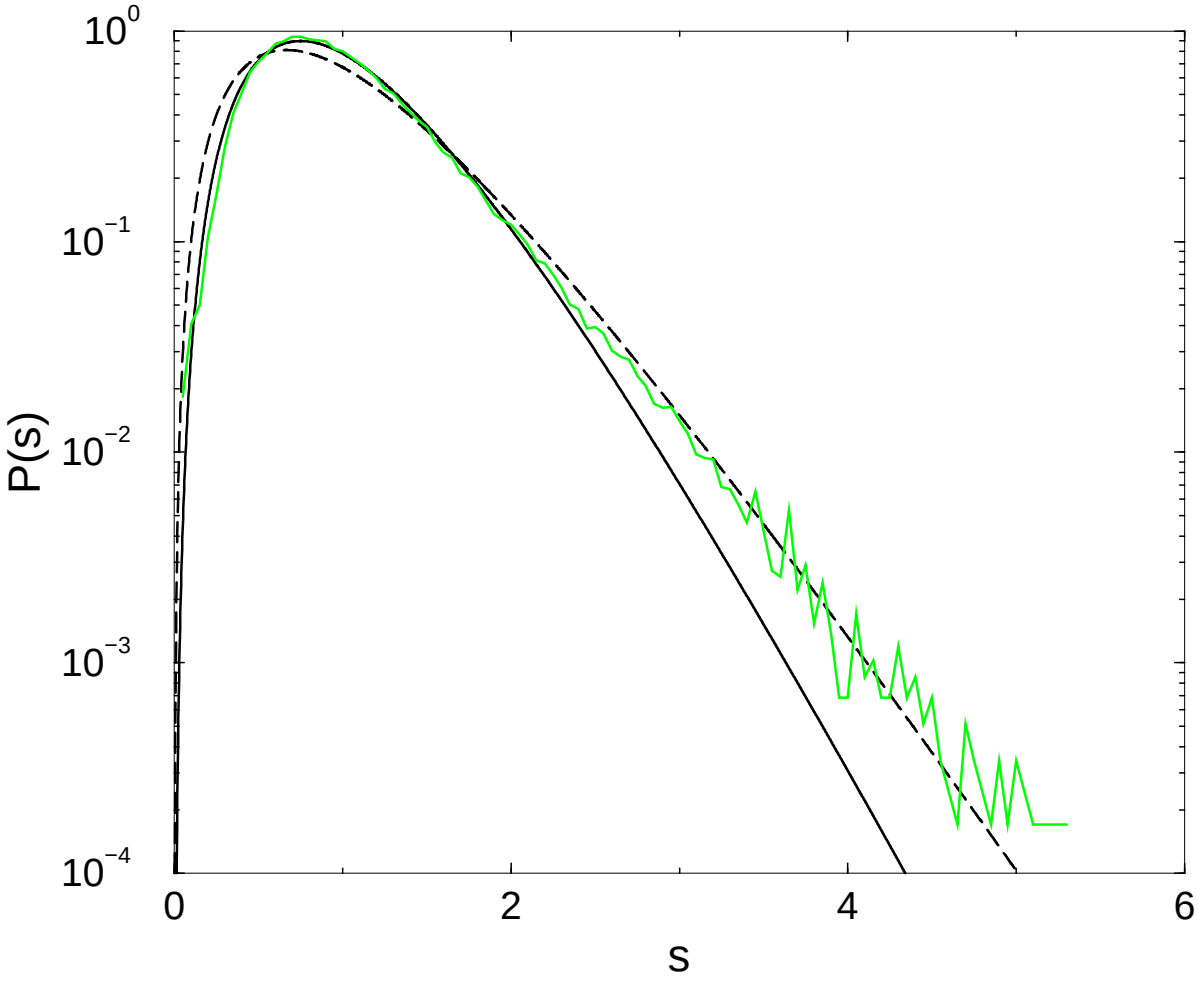

FIG. 9: Unfolded distribution for a randomly shuffled sequence of votes for the PRI presidential candidate (green line). A daisy model of rank $r = 2$ and 3 (see text for details) are shown in dashed and continuous black line.

V. ACKNOWLEDGMENTS

This work was supported by DGAPA-UNAM project IN-104400. We want to thank C. Badillo useful comments. We wish to thank to E. Morfin for his invaluable help with the databasis of the PREP and to A. Baqueiro for giving us his data of capture in real time of the PREP for Fig. 1.

[1] Borghesi C, Galam S. Chaotic, Physical Review E 73 (6), 066118 (2006).

[2] Miller C.A., Social Studies of Science 34 (4), 501-530 (2004).

[3] All data were obtained form the web page of the IFE, http://www.ife.org.mx/prep2006/bd_prep2006/bd_prep2006.htm

[4] The meaning of PAN, PBT, and APM are *Partido Acción Nacional, Alianza por el bien de todos y alianza por México*, respectively. The PBT is composed by the following parties: *Partido de la Revolución Democrática, Partido del Trabajo* and *Convergencia*. The coalition APM is composed by the *Partido Revolucionario Institucional, PRI* and by the *Partido Verde Ecologísta*. In our appreciation this coalition is right-center although there are different opinions. The PAN is now the party in the federal government while the PRI was in the same position during more than 70 years.

[5] http://em.fis.unam.mx/public/mochan/elecciones/indexen.html.

[6] Jiao Y, Syau YR, Lee ES. Fuzzy adaptive network in presidential elections, Mathematical and Computer Modelling 43 (3-4), 244-253 (2006).

[7] Lang J. Some representation and computational issues in social choice. Symbolic and Quantitative Approaches to Reasoning With Uncertainty, Proceedings Lecture Notes in Computers Science 3571, 15-26 (2005).

[8] Merlin V, Valognes F. The impact of indifferent voters on the likelihood of some voting paradoxes. Mathematical Social Sciences 48 (3), 343-361 (2004).

[9] R.N. Costa Filho *et al* Phys. Rev E. **60**, 1067 (1999); Physica A **322**, 698 (2003).

[10] A.A. Moreira, D.R. Paula, R.N. Costa Filho and J.S. Andrade Jr. Phys. Rev E. **73**, 065101(R) (2006).

[11] G. Baez, H. Hernández Saldaña and R.A. Méndez-Sánchez. "Power-laws in the mexican elections of 2006". In progress.

[12] H. Hernandez-Saldaña, J. Flores and T.H. Seligman. Phys. Rev E. **60**, 449 (1999).